Los Alamos
NATIONAL LABORATORY
EST. 1943

LA-UR-21-24987
Approved for public release; distribution is unlimited.

Title: (U) Archiving Physics Data for Today and Tomorrow

Author(s): Ham, Michael Isaiah

Intended for: Physics Cafe Presentation

Issued: 2021-05-24

Disclaimer:
Los Alamos National Laboratory, an affirmative action/equal opportunity employer, is operated by Triad National Security, LLC for the National Nuclear Security Administration of U.S. Department of Energy under contract 89233218CNA000001. By approving this article, the publisher recognizes that the U.S. Government retains nonexclusive, royalty-free license to publish or reproduce the published form of this contribution, or to allow others to do so, for U.S. Government purposes. Los Alamos National Laboratory requests that the publisher identify this article as work performed under the auspices of the U.S. Department of Energy. Los Alamos National Laboratory strongly supports academic freedom and a researcher's right to publish; as an institution, however, the Laboratory does not endorse the viewpoint of a publication or guarantee its technical correctness.

Welcome to Physics Café

Mike Ham, P-2:

(U) Archiving Physics Data for Today and Tomorrow

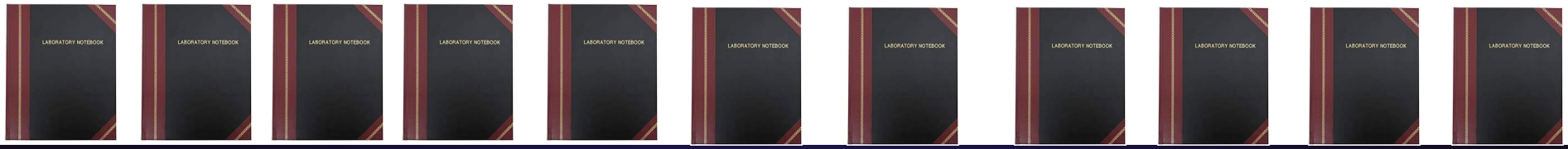
April 15, 2021

My Data Background

Project	Types of Data Collected / Accessible to community	Status of Data
Legos	Mental / No	Still in use
UNT Liquid Nitrogen Car	Notebooks / No	Discarded during a move
UNT Accelerator Laboratory	Notebooks / No	Publications
UNT Neurophysiology Laboratory / LANL T Division	1. Notebooks / No 2. Electronic Database / No 3. Electronic data files / No 4. Electronic Documents / Only Published	1. Publications 2. Removable hard drives that maybe work (13 years)
LANL Computer Vision	1. Notebooks / Yes 2. Myriad of computer programs and data files / No 3. Electronic Documents / Only Published	1. Notebooks in my office 2. Computer programs may exist on a thumb drive 3. Publications 4. Data likely inaccessible (computer lockout)
LANL Reaction History	1. Notebooks / Yes 2. Electronic Documents / Yes 3. Scanned Repository / Yes	1. X Online Vault 2. Red Repository 3. SSDB 4. Common Modeling Framework

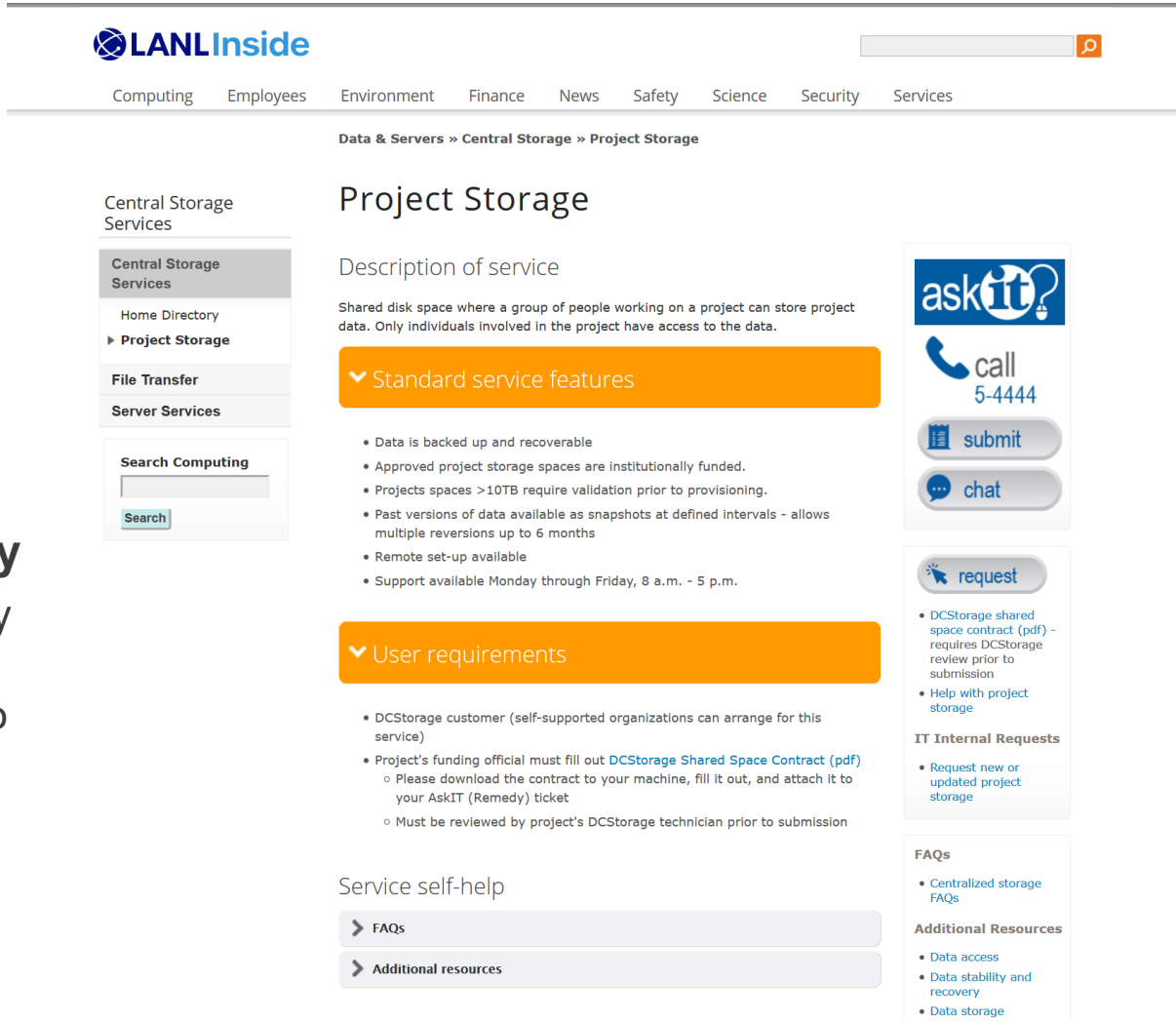
LANL did/does some things well with data

- **Most high level paper documents from historic shots were preserved.**
  - Most results data were preserved (films)
  - We have an active archival effort (WRS)
  - Culture of recording data
- **Culture has reverence for data when answering important problems**
- **When I got to LANL we were told to preserve notebooks**
  - Notebooks are provided! (your own personal repository)
  - *Have saved multiple experiments, use a notebook*
  - Probably mostly useless for others to read, but that's okay



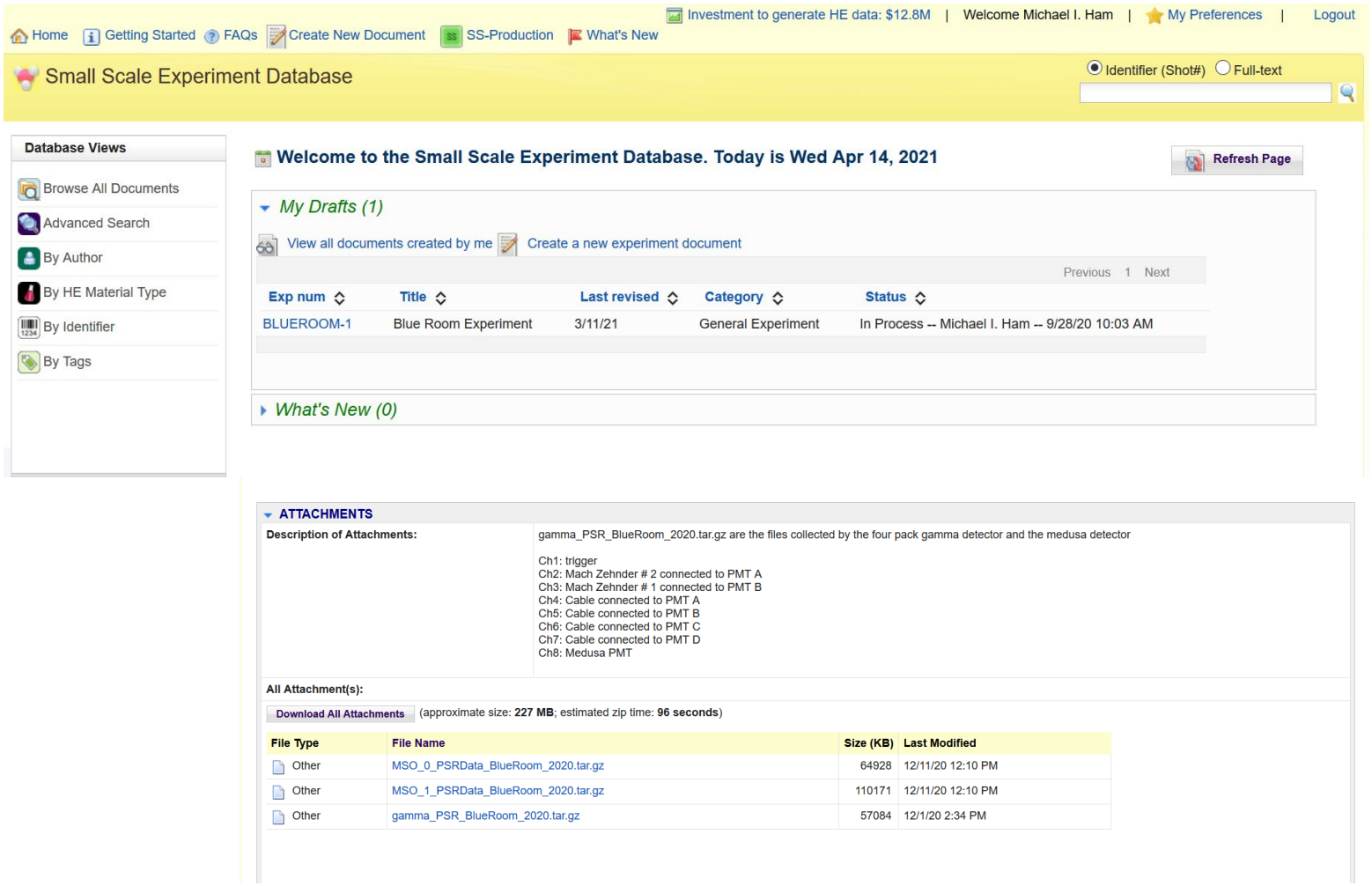
LANL does some things badly with data

- **Some historic data was deemed unimportant and discarded after testing**
  - Cables
- **Archival efforts tend to be extremely ad-hoc**
- **Institutional data repositories are not well known and not user friendly**
  - No institutionally accepted way to archive your digital data for use by the laboratory, it's up to you and your team/group/division.
    - That has some advantages (flexibility)
    - But also disadvantages in that the data are everywhere and nowhere



https://int.lanl.gov/computing/data-servers/central-storage/storage-types.shtml

Making data available to the future



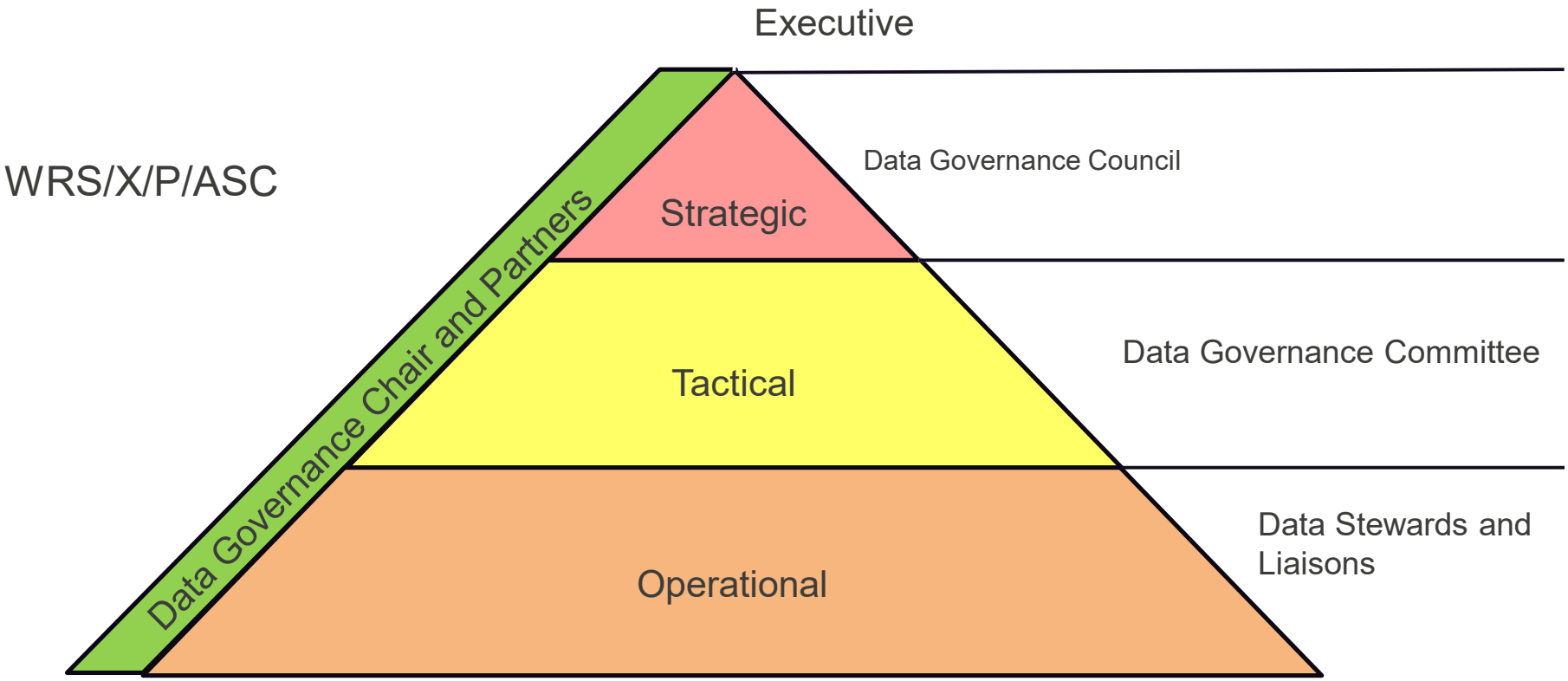
http://smallscale-y.lanl.gov

What am I Currently doing?

- **Assisting in updating internal digital archives**
- **Storing Data in the Small Scale Data Base (SSDB)**
  - Blue Room data (yellow)
  - MIE scattering data (yellow)
  - UGT data (red)
  - Surface Imaging Data?
- **Working directly with X-division to put data with pedigree into the Common Modelling Framework (CMF)**
  - Not complete agreement that this is where it should go, but it does work and is simple to implement.
- **Assisting in Establishing Data Governance Body with WRS**

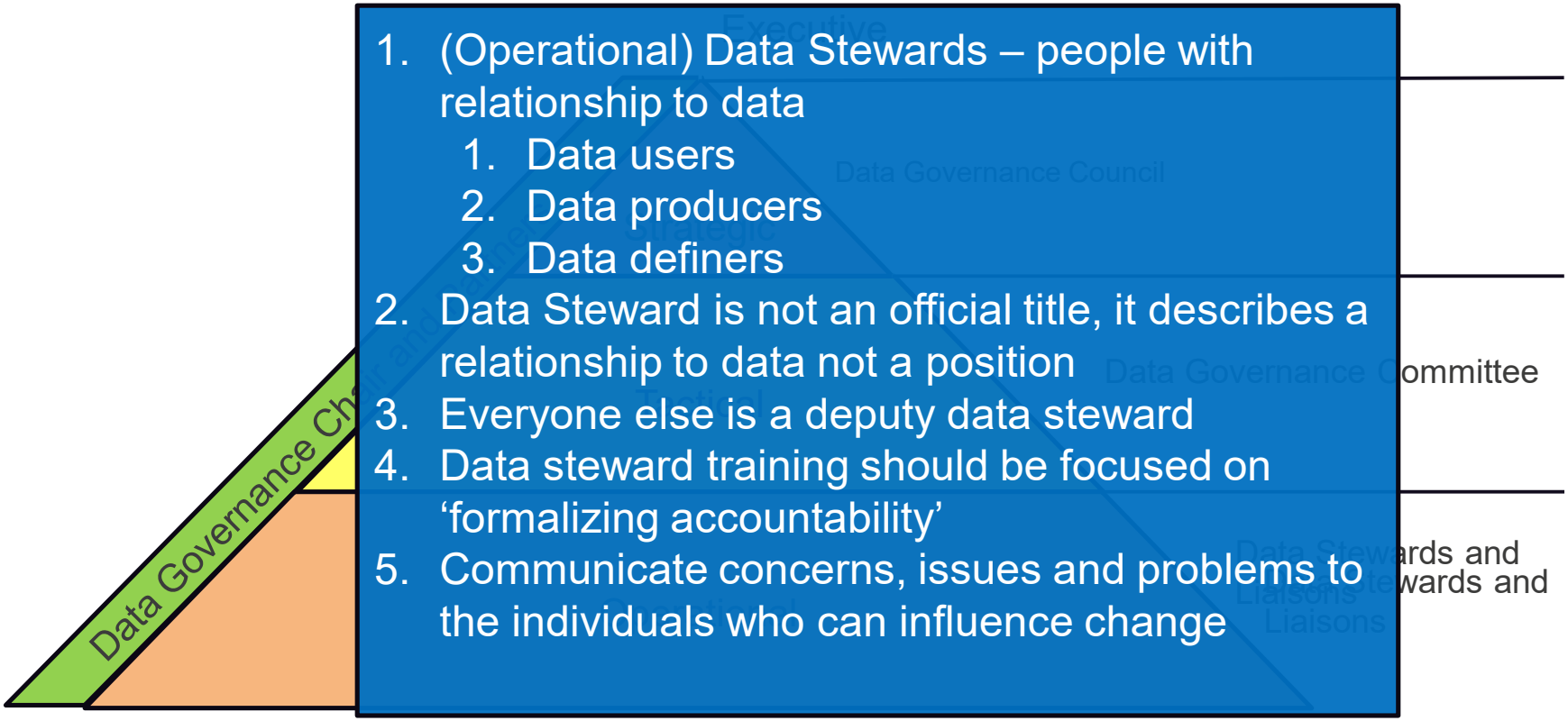
Data Governance Body

Roles and Responsibilities



Data Governance Body

Roles and Responsibilities



1. (Operational) Data Stewards – people with relationship to data
    1. Data users
    2. Data producers
    3. Data definers
2. Data Steward is not an official title, it describes a relationship to data not a position
3. Everyone else is a deputy data steward
4. Data steward training should be focused on 'formalizing accountability'
5. Communicate concerns, issues and problems to the individuals who can influence change